OSU Dept. of Computer Science and Engineering Technical Report OSU-CISRC-03/05-TR18.

Modeling Pedestrian Behavior and Detecting Event Anomalies using a Seasonal Kalman Filter

James W. Davis Mark A. Keck
Dept. of Computer Science and Engineering
Ohio State University
Columbus OH 43210 USA
{jwdavis,keck}@cse.ohio-state.edu

Abstract

We present a seasonal state-space model using Kalman recursions to learn and predict structured behavior patterns. The model is employed to detect events using the learned expectations of typical scene activity. We demonstrate the approach for modeling the expected number of pedestrians in a scene over hour-long periods (over multiple days) and for detecting event anomalies. The framework provides a single long-term model by exploiting the natural seasonal trends in daily human activity.

1. Introduction

An important aspect of any intelligent surveillance and monitoring system, beyond person detection and activity recognition, is the ability to automatically learn/discover patterns or trends of scene behavior and to identify anomalous events. For example, one may wish to be alerted when a person is present in a secure unpopulated area, if a person is seen at a particular time of day when there should be no people present (e.g., late at night), or if an individual remains in an area for an extended time. We also may be interested in more dynamic events such as if a person is seen following an unusual path or running across a normally quiescent area.

As most video cameras are positioned to monitor a fixed area over long periods of time (even several years), it is not unreasonable to insist that the system should *learn the activity trends by observation* and *adapt itself through experience* (modifying its model parameters). Obviously, certain types of event alerts will require manually-created detectors, but statistical modeling of activity trends over long periods can be employed to automatically derive expectations of scene behaviors. Then, any observed behavior that violates the scene expectation can be considered to be an anomalous event. The main advantage of a statistical approach is that individual detectors for all possible atypical events do not need to be hand-crafted. Specifics of the event (e.g., location, action, time-of-day) could also be used to classify the anomaly for video storage and retrieval purposes (particularly helpful to a video analyst).

Due to our highly scheduled lifestyles, human activity tends to exhibit the pervasive phenomenon of "seasonality", referring to the tendency to repeat a high-and-low occurrence patterns of behavior across the day/week. For example, on a college campus many people will be seen *between* class periods and few *during* the class times (except for the occasional latecomers). In Fig. 1, we show a plot of the number of people present in a particular scene on a campus from 9-10am (sampled every 15 seconds) on three Mondays. Though there is a fairly large variation at any particular time across the three days, there does exist an overall rise-and-fall pattern for the presence of the people across the hour (the peak occurs at the time between classes, 9:20-9:30am). Such time-structured patterns are common in both public and private spaces. A method that can exploit natural seasonal activity trends will be better-suited to represent and predict the behavior patterns for anomaly detection.

In this paper, we present a seasonal state-space model for learning, predicting, and validating expectations of *typical* scene activity. In particular, we focus on modeling the presence patterns (counts) of pedestrians in hour-long time periods on particular days of the week on a college campus. The approach is based on a seasonal time series model applied over hourly periods, rather than using a brute-force method with a separate model for *each* second of *each* day. The state and noise parameters for the seasonal model are initialized from training data, and Kalman recursions are employed to optimally estimate the state sequence and to forecast the next time period. We detect an anomalous event when a new observation exceeds the confidence range for the predicted (expected) behavior. The model is updated with new observations to adapt to slowly changing trends and seasonality. We demonstrate the approach on synthetic data and thermal video surveillance imagery.

The remainder of this paper is described as follows. We begin with a review of related work (Sect. 2). Next we

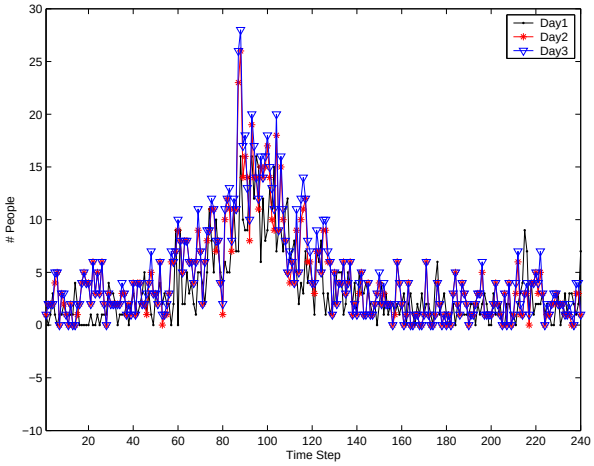

Figure 1: Number of people manually counted in a particular scene from 9-10am (every 15 sec.) on three Mondays.

describe the background and formulation for the proposed seasonal framework (Sect. 3). We then present experimental results (Sect. 4). Lastly, we conclude with a summary of the research and describe future work (Sect. 5).

2. Related Work

To model activity patterns and detect anomalous events, several graph-based and trajectory-based approaches have been proposed. A Dynamic Bayesian Network was employed in [10] to model temporally-correlated events for detecting typical and atypical behaviors. In [8], the trajectories of people were used to build a semantic scene model (entry zones, paths, junctions) from which activity expectations could be extracted. In [6], a statistical model of pedestrian trajectories was used to learn typical paths and to identify incidents of unexpected behavior. A polygonal shape configuration and its deformation for path behaviors of people were examined for abnormal changes in [9]. Anomalous office activities were detected in [1] by identifying inputs having a low likelihood to an entropically-estimated HMM.

Aggregate trajectory-based methods can be used to learn interesting path models, but these methods generally ignore how the occurrence of these behaviors change with respect to time (e.g., over hours or days). Our proposed approach exploits the natural seasonality of behavior to learn proper time-based expectations for detecting event anomalies.

3. Seasonal Forecasting

The classical decomposition model for a seasonal time series is an explicit representation composed of the underlying *trend*, *seasonal variation*, and *irregular (random) noise* components [2]. Other more generic techniques include autoregressive moving average models (ARMA, ARIMA, SARIMA). An advantage to the classical decomposition model is that we have a representation for an underlying process model of the time series.

The classical decomposition model for a univariate observation y at time t is given by

$$y_t = m_t + s_t + n_t \tag{1}$$

where m_t is the trend component, s_t is the seasonal component, and n_t is noise. The seasonal component s_t has a period of d with the properties $s_{t+d} = s_t$ and $\sum_{i=0}^{d-1} s_{t+i} = 0$ (i.e., zero mean). In Fig. 1, the trend m_t is a constant value (~ 4) with the seasonal values having a unimodal rise-and-fall pattern over the period (with the addition of noise). Once the three components have been estimated (from training data), the properties of the time series can be used to predict a future observation y_{t+n} (n-step prediction) by running the model forward in time. This will become useful for detecting anomalies.

3.1. Seasonal Kalman Filter (SKF)

A *state-space representation* of the classical decomposition model can be used to avoid the deterministic strictness of the components by allowing the trend and seasonal components to evolve randomly in a recursive manner [2]. A state-space model for a time series $\mathbf{Y}_t$ (potentially multivariate) consists of two fundamental equations:

$$\mathbf{Y}_t = G\mathbf{X}_t + \mathbf{W}_t \tag{2}$$
$$\mathbf{X}_{t+1} = H\mathbf{X}_t + \mathbf{V}_t \tag{3}$$

The *observation equation* (Eqn. 2) gives $\mathbf{Y}_t$ as a linear function of the state variable $\mathbf{X}_t$ plus Gaussian measurement/observation noise $\mathbf{W}_t$, and the *state equation* (Eqn. 3) determines the next state $\mathbf{X}_{t+1}$ from a linear function of the current state $\mathbf{X}_t$ plus Gaussian process noise $\mathbf{V}_t$. Typically, the noise is treated as independent of time (dropping the subscript t).

The univariate decomposition model (Eqn. 1) can be formulated in a recursive state-space model (here with $\mathbf{Y}_t = y_t$) as follows. First, the d-dimensional state $\mathbf{X}_t$ is set to

$$\mathbf{X}_t = \begin{bmatrix} m_t & s_t & s_{t-1} & \cdots & s_{t-d+2} \end{bmatrix}^\top \tag{4}$$

consisting of the trend and the $(d-1)$ most-recent seasonality values. The corresponding observation and state equations are given by

$$G = \begin{bmatrix} 1 & 1 & \cdots & 0 \end{bmatrix} \tag{5}$$
$$\mathbf{W} = \begin{bmatrix} \mathcal{N}(\sigma_w) \end{bmatrix} \tag{6}$$

$$H = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & -1 & \cdots & -1 & -1 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \quad (7)$$

$$\mathbf{V} = \begin{bmatrix} \mathcal{N}(\sigma_{v1}) & \mathcal{N}(\sigma_{v2}) & 0 & \cdots & 0 \end{bmatrix}^\top \quad (8)$$

For the observation equation (using G and $\mathbf{W}$), the matrix G adds together the trend m_t and the current seasonality value s_t, followed by the inclusion of the observation noise in $\mathbf{W}$. For the state equation (using H and $\mathbf{V}$), the matrix H (when applied to $\mathbf{X}_t$) simply retains the trend m_t, but updates the seasonality s_t using the zero-mean constraint with the most-recent $(d-1)$ seasonality values; the most-recent $(d-2)$ seasonality values are shifted along through time. We only add process noise ($\mathbf{V}$) to the two initial state elements (m_t, s_t), as the remaining terms are only shifted seasonality values.

Given values for the initial state $\mathbf{X}_0$, the noise (co)variances, and a sequence of observations $\mathbf{Y}_t^{obs}$, standard Kalman recursions can be employed to optimally estimate the $\mathbf{X}_t$ state sequence [7]. From the state sequence, the predicted observations are given by $\mathbf{Y}_t^{pred} = G\mathbf{X}_t$. As the observed data are continually incorporated into the model over time (using either Kalman filtering or smoothing recursions), the model adapts to slowly changing trends and seasonality.

3.2. Anomaly Detection

The Kalman recursions for the state-space model also propagate an error covariance P_t for each state $\mathbf{X}_t$. When we encounter a new observation, we can use the estimated state error covariance as a means to set the confidence levels/bounds (e.g., 90% confidence) for detecting outliers (events) from the predicted observation.

To transform the state error covariance P_t into a (co)variance measurement in the observation space, we use

$$C_t = GP_tG^\top + R \quad (9)$$

with the observation noise covariance $R = E[\mathbf{W}\mathbf{W}^\top]$. We can then use the Mahalanobis distance to determine the statistical match of the actual observation $\mathbf{Y}_t^{obs}$ to the model prediction $\mathbf{Y}_t^{pred}$. The general (multivariate) Mahalanobis distance is given by

$$D = \sqrt{(\mathbf{Y}_t^{obs} - \mathbf{Y}_t^{pred})^\top C_t^{-1} (\mathbf{Y}_t^{obs} - \mathbf{Y}_t^{pred})} \quad (10)$$

The distance D is given in standard deviations, therefore we can threshold the distance with statistical confidence (e.g., >3 SD is a statistical outlier).

The distance calculation could be applied to each new observation $\mathbf{Y}_t^{obs}$ to detect anomalous events, followed by an immediate update of the model (given the new observation). However, since we are employing seasonality constraints within the model, we instead employ a seasonal forecast of the *entire* period and use the state error covariance P computed for the first time step for the entire period. We only update the model and P after the entire period has been tested.

This method is used for two reasons. First, the model update with Kalman recursions (including matrix inversions) is computationally taxing when the length of the period is large. For the data shown in Fig. 1, the period is 240 time steps. Thus $\mathbf{X}_t$ is of size 240×1 (H and G are similarly large). In our approach, we forecast the entire period before the period begins, which can be computed on an auxiliary processor. Second, if we blindly incorporate severe outliers into the model as they are encountered (true anomalies beyond the noise expectations), the model can be altered and the remaining predictions for the period can be adversely affected. We instead incorporate the observations into the model only after the entire period has been examined, and also limit the influence of any outliers to ± 3 SD of the expected value (to allow only slow changes over time).

3.3. Extension for SKF Modeling of Pedestrian Presence

The above univariate SKF formulation is well-suited to seasonal patterns, however there is a problem if it is applied directly to the particular class of seasonal data as shown in Fig. 1. In that data, there is a fairly large variation of pedestrian counts at each time step (across the three days). Thus, after initialization and training, a new sequence of observations need only be *individually* within the wide tolerance at each time step to be considered "typical" and thus no strict consideration of the overall pattern for the period is enforced. This would certainly cause a problem if we took the brute-force approach of using an individual Kalman filter for each time-step (for each day).

Consider the case for a new period of observations with a low and constant pedestrian count of 0-3 people. In relation to Fig. 1, at best only a few outliers near the bump in the time period will be detected as atypical. Individually, most observations may be within the wide bounds of the model, but *collectively* there is no rise-and-fall seasonality as expected. One may be tempted to first smooth the pedestrian counts to reduce this variation and then model these smooth patterns with the SKF. But unfortunately this would smooth-out and ignore any brief, yet meaningful, spikes or vacancies in pedestrian counts. Such events do in fact occur and could be especially prevalent in time-lapse monitoring.

Our approach to handle this problem is to employ two SKF models to account for the raw and locally-smoothed

pedestrian counts. Thus we can better model the seasonality with the locally-smoothed counts yet still detect spikes/vacancies using the raw counts. If we employed a simple linear smoothing filter (e.g., mean/average), we could easily incorporate it into a single unified SKF model, rather than employing two independent models. However, such a linear smoothing filter will be adversely affected by event spikes (corrupting values in the local area around the events). Due to the brief and prominent nature of outliers/spikes, we found that the use of a median filter is more appropriate than a mean or Gaussian. The median will not be corrupted by the brief event spikes. In our domain of pedestrian counts (sampled every 15 sec.), we found that a 12-tap causal median filter was appropriate to filter out short event bursts (it typically takes less than 2 minutes for people to traverse the scene).

Our new method is simply to run a raw-input SKF in parallel with a median-filtered SFK. If an outlier is found (using Eqn. 10) in either of the models, then an event is registered. Hence, we can now detect precise event spikes using the raw data (without corrupting surrounding time steps) and ensure longer-term behavior patterns for the period (e.g., general rise-and-fall patterns) using the median data.

4. Experiments

To evaluate the proposed SKF framework, we tested the approach for modeling pedestrian counts and detecting anomalies using synthetic data and real thermal imagery with a pedestrian detector algorithm.

4.0.1. Model Initialization

We begin with an initialization of the state estimate $\mathbf{X}_0$ and the variances for the noises $\mathbf{W}$ and $\mathbf{V}$. To give a reasonable estimate of $\mathbf{X}_0$, we calculate the mean for the first period of the training data (detector results) and assign it to m_t, and then use the most recent $(d-1)$ mean-subtracted data values as the seasonality estimates for s_{t-d+2} through s_t. The initial state error is set large to $P_0 = 10^5 I$ and is reduced during training.

To estimate the noise parameters (from the real data), we compute the measurement noise (co)variance $R = E[\mathbf{W}\mathbf{W}^\top]$ from training data using the differences between the manually-counted people and the automatic detector results per frame (for all training images). This is done independently for each model (raw, median). Similarly, the process noise covariance $Q = E[\mathbf{V}\mathbf{V}^\top]$ is estimated from training data for each model (raw, median) using the period-long mean trend differences and the differences between the mean-subtracted count differences at each time step (across multiple periods). Here we use the manual counts to estimate the process noise, as they are not corrupted by measurement noise. We note that an alternate maximum likelihood noise estimation process could also be employed [2].

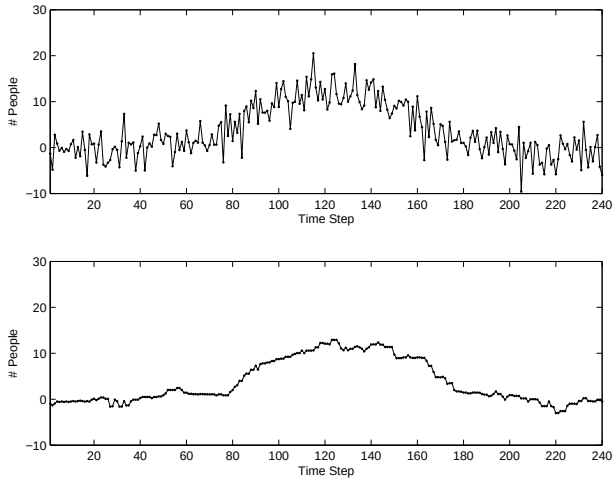

Figure 2: One period of synthetic data (raw and median-filtered values).

4.1. Synthetic Data

We created a set of synthetic data similar to the sequences shown in Fig. 1. We first employed a Gaussian rise-and-fall pattern stretched over 240 samples ($\sigma = 24$). Next the pattern was shifted up by a constant trend value (~ 4). To incorporate the noise, we began by substituting Eqn. 3 into Eqn. 2 (assuming $\mathbf{W}$ and $\mathbf{V}$ are uncorrelated) to yield

$$\mathbf{Y}_t = GH\mathbf{X}_{t-1} + \omega \tag{11}$$

where $\omega = G\mathbf{V} + \mathbf{W}$ and $\Omega = E[\omega\omega^\top] = GQG^\top + R$. We used the Q and R (co)variances estimated from three Monday training days using the method provided above in Sect. 4.0.1. The noise ω was added to the raw synthetic data and the corresponding median data was formed using a 12-tap causal median filter. An example of one synthetic period (raw, median) is presented in Fig. 2.

We trained the SKF models using 5 synthetic periods and then analyzed the model on new synthetic test data. First we tested a new period generated using the same parameters employed to create the synthetic training data. As expected the model closely matched the new test data. The largest deviation found (using Eqn. 10) in the test period was 2.7 SD. Next we applied 4 event spikes (using a constant ± 10 SD of the noise ω) to a new test period, as shown in Fig. 3 (top plot). Using a threshold of 3 SD in Eqn. 10, the 4 event spikes (and only those 4) were detected as anomalies. The median-filtered values were unaffected (as desired) by the events, and thus the events were detected solely in the raw count model. Lastly, we tested a uniform pedestrian count

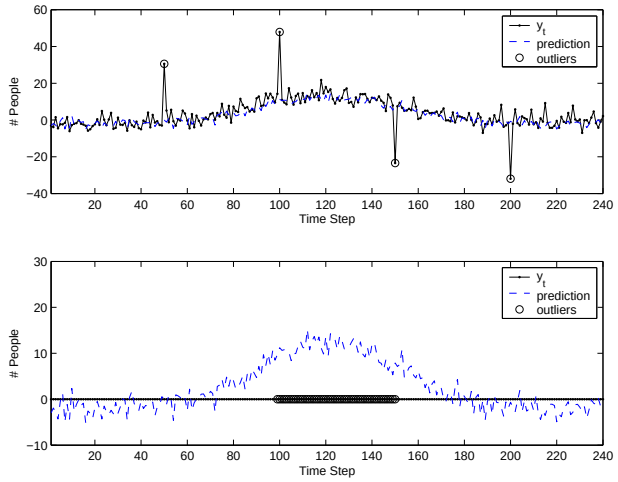

Figure 3: Synthetic event results. Top plot: Event spikes. Bottom plot: Inactivity. Events are marked with 'o'.

of 0 (inactivity) over the 240 samples. The results are shown in Fig. 3 (bottom plot), in which 52 samples were identified as outliers (using the same 3 SD threshold as before). In this example, all of the outliers were detected in the median model, but a few (4) of the events were also found in the raw model.

We also added several normal periods after the inactivity period to determine how long the model would take to recover from the outliers after using the constrained update process (as explained at the end of Sect. 3.2). The model required 2 periods of normal activity (after the inactivity period) to stabilize and report absolutely no events.

4.2. Thermal Surveillance Imagery

To test the approach with real data, we recorded several thermal video sequences of pedestrian traffic on a college campus from 9-10am (1 frame captured every 15 sec.) on several Mondays, Wednesdays, and Fridays at a particular location (15 days, 3600 frames total). The number of pedestrians in each image were manually counted for ground truth comparison. The manual counts for three Mondays are shown in Fig. 1.

To automatically detect the pedestrians in each image, we employed the two-stage template-based method of [3]. The approach initially performs a fast screening procedure using a generalized thermal contour template [5, 4] to locate potential pedestrian locations. Next an AdaBoosted ensemble classifier using automatically tuned filters/classifiers is employed to test the hypothesized pedestrian locations. This method was trained using a sample of 114 frames from the thermal video collection, and then the detector was used to count the number of pedestrians in all of the images.

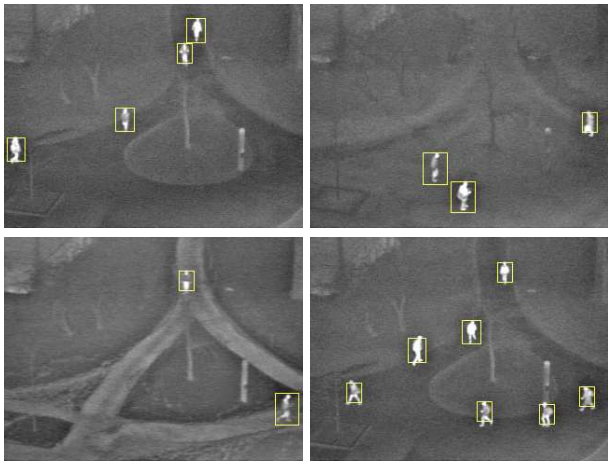

Figure 4: Pedestrian detection results in thermal imagery.

Example thermal images showing the detection results are shown in Fig. 4.

We assigned a separate raw/median SKF model to each day of the week (e.g., only for Mondays). Though the overall pattern in each day of the week was somewhat similar, we found that Fridays during that time period were different enough from Mondays and Wednesdays to warrant a separate model (fewer people were present in the scene on Fridays). We trained the models (for each day of the week) using 3 manually-counted days. The noise covariances Q and R were estimated from the training data, and a 12-tap causal median filter was used for smoothing. To set the detection threshold for a raw (or median) model, we tested each training day in parallel and selected an outlier threshold such that no events occurred on any of its training data. The event thresholds (raw/median) were 2.0/1.5, 4.5/1.0, and 3.0/1.0 SD for the Monday, Wednesday, and Friday models, respectively.

We show the testing results on two new days for each model in Fig. 5. As described in Sect. 3.2, the models are updated after each period. In the Monday sequence results, 1 anomaly was detected in the median-filtered data (the first event) and 3 anomalies were found in the raw counts. We show the raw data with the events labeled in the top plot of Fig. 5. Upon inspection of the data, we feel these detections were reasonable. For the Wednesday results, we found a short cluster of detections (12 time steps long) only in the median model. In the middle plot of Fig. 5, we show the median data with the detected outliers. It is clear that there was a significantly lower presence of people during that short interval. Lastly, in the bottom plot of Fig. 5, we detected no events in the Friday sequences.

We additionally were able to record a Monday on a holiday showing few pedestrians in the scene. The detected

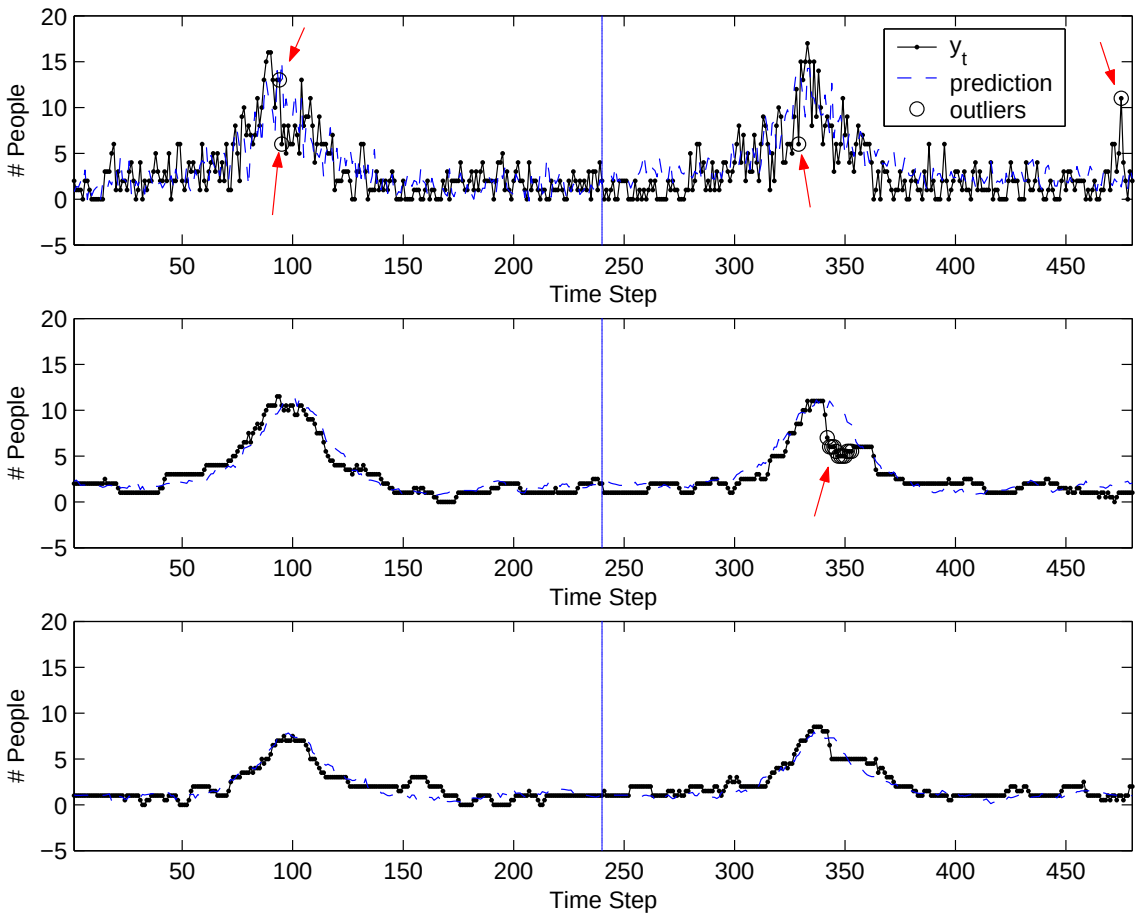

Figure 5: Testing results for 2 periods of real data. Top plot: Mondays (raw). Middle plot: Wednesdays (median). Bottom plot: Fridays (median).

events (61 anomalies) are shown in Fig. 6 (top plot). Events were found in both the raw and median models (the median model detected every event). Since we had no day available with truly known event spikes, we simulated a large pedestrian count near the end of the period (for 1 min.) for one of the Monday training days (simulating a tour group passing through the scene). The detection result targeting the event is shown in Fig. 6 (bottom plot). As in the previous synthetic case, the model was able to target the events specifically without giving false positives for surrounding valid presence counts.

5. Summary and Conclusions

We presented a seasonal state-based model for learning, predicting, and validating longer-term behavioral trends. The approach is based on a model containing the trend, seasonal variation, and noise over a time period (e.g., hour), rather than employing a brute-force method with a separate model for each second of each day. Kalman recursions are employed to optimally estimate the state sequence for given set of observations. Predictions for the following period and their confidences are used to detect deviations (event anomalies) from the expected behavior.

We demonstrated the framework for the task of modeling the presence patterns of pedestrians on a campus during particular days of the week. We initially showed the performance of the model on synthetic examples. Then we demonstrated the model on actual pedestrian counts provided by a person detection algorithm applied to a collection of thermal surveillance imagery. The results showed that the model is well-grounded in the seasonal nature of the data and that it is capable of reliably detecting events.

In future work, we will extend the approach to include action recognition labels in a multivariate prediction of seasonal presence and action. We will also collect a video database of hour-long periods during the day and across months to evaluate the longer-term applicability of the ap-

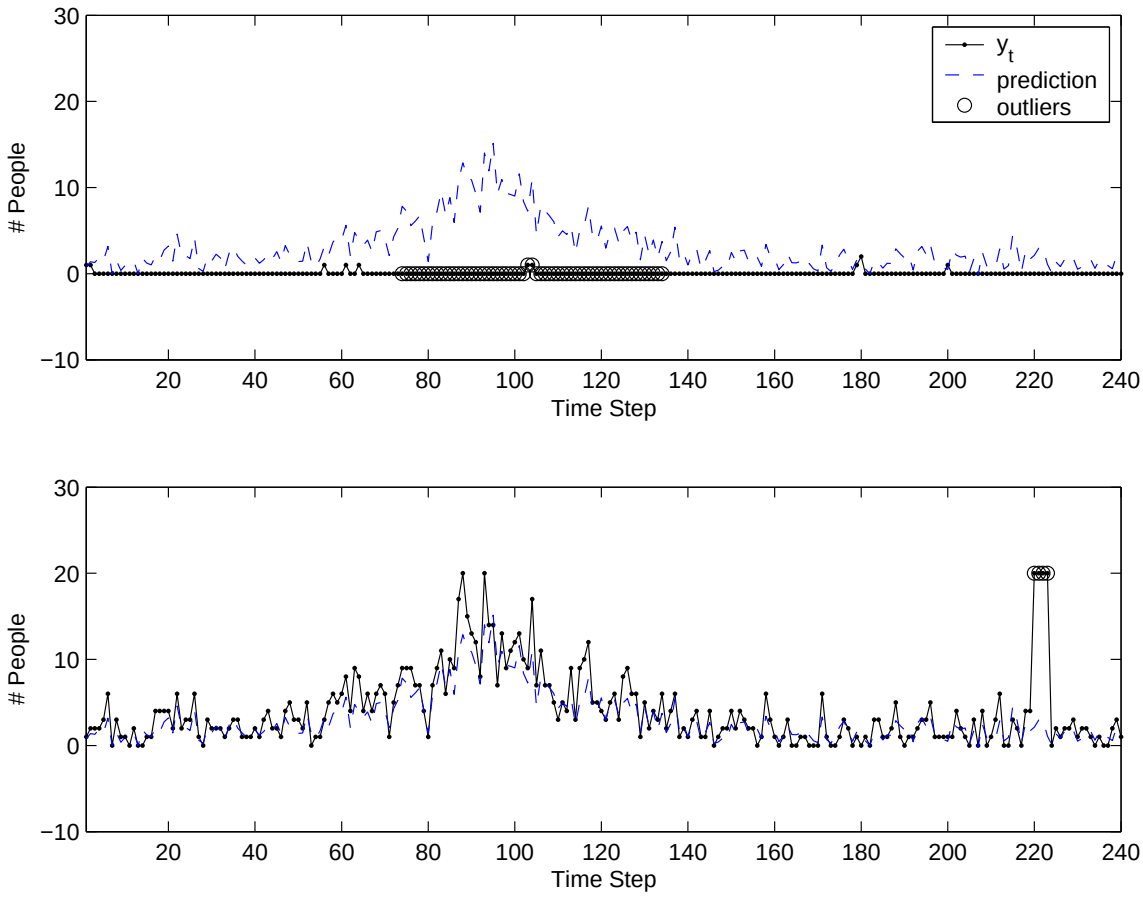

Figure 6: Anomalies. Top plot: Holiday inactivity. Bottom plot: Simulated event spikes.

proach. Furthermore, we plan to use prior event knowledge, such as a calendar with scheduled holidays, to further constrain the model updates when anomalies are encountered. As seasonal behavior patterns are quite common in our society, we expect this model to be useful for further study in relation to automatic surveillance and monitoring systems.

Acknowledgments

This research was supported in part by the National Science Foundation under grant No. 0236653 and the U.S. Army Night Vision Laboratory.

References

[1] M. Brand and V. Kettnaker. Discovery and segmenation of activities in video. *IEEE Trans. Patt. Analy. and Mach. Intell.*, 22(8):844–851, 2000.

[2] P. Brockwell and R. Davis. *Introduction to Time Series and Forecasting*. Springer-Verlag, New York, 2002.

[3] J. Davis and M. Keck. A two-stage template approach to person detection in thermal imagery. In *Proc. Wkshp. Applications of Comp. Vis.*, 2005.

[4] J. Davis and V. Sharma. Robust background-subtraction for person detection in thermal imagery. In *IEEE Int. Wkshp. on Object Tracking and Classification Beyond the Visible Spectrum*, 2004.

[5] J. Davis and V. Sharma. Robust detection of people in thermal imagery. In *Proc. Int. Conf. Pat. Rec.*, pages 713–716, 2004.

[6] N. Johnson and D. Hogg. Learning the distribution of object trajectories for event recognition. In *Brit. Mach. Vis. Conf.*, pages 583–592, 1995.

[7] R. Kalman. A new approach to linear filtering and prediction problems. *Trans. of the ASME-Journal of Basic Engineering*, 82 (Series D):35–45, 1960.

[8] D. Makris and T. Ellis. Automatic learning of an activity-based semantic scene model. In *Advanced Video and Signal Based Surveillance*, pages 183–188. IEEE, 2003.

[9] N. Vaswani, A. Chowdhury, and R. Chellappa. Activity recognition using the dynamics of the configuration of interacting objects. In *Proc. Comp. Vis. and Pattern Rec.*, pages 633–640, 2003.

[10] T. Xiang and S. Gong. Discovering Bayesian causality among visual events in a complex outdoor scene. In *Advanced Video and Signal Based Surveillance*, pages 177–182. IEEE, 2003.